Lecture 26 : Hill-Robertson interference

MATH285K - Spring 2010 *Lecturer: Sebastien Roch*

References: [Dur08, Chapter 8].

1 K alleles

When considering more than 2 alleles, we need a vector of frequencies to keep track of the state of the population. In the diffusion limit, this leads to a multidimensional diffusion. Although the limit process can be derived from the Wright-Fisher, as an aside we describe instead a derivation based on the so-called *Moran model*.

1.1 Moran model

The Moran model is a population dynamics model similar to the Wright-Fisher model where the assumption of non-overlapping generations is relaxed. We first describe the model without mutation or selection.

No mutation/selection. Suppose we have a population with $2N$ haploids. In continuous time, each individual dies at rate 1 and is replaced by a copy of a uniformly selected individual in the population.

We observe that, when run backwards in time at rate N, this process leads to Kingman's coalescent in the $N \to \infty$ limit. Indeed, with k lineages, a death occurs at rate kN and a replacement within the current samples occurs with probability

$$\frac{k-1}{2N},$$

leading to a coalescence at rate

$$N\frac{k(k-1)}{2N} \to \frac{k(k-1)}{2},$$

as in the standard coalescent.

Including mutation/selection. Now assume that each individual has one of K alleles. Mutations between states i and j occur at rate ϕ_{ij}. Denote by $n = (n_1, \ldots, n_K)$ the state of the population. Moreover, letting the relative fitness of allele i be $1 - s_i$ for $i = 1, \ldots, K$, we assume that overall rate of change is

$$n \to n + e_i - e_j,$$

at rate

$$n_j \left(\frac{n_i}{2N}(1 - s_i) + \phi_{ji} \right).$$

In other words, when a j-individual dies we propose a replacement uniformly at random. Say an i-individual is picked for the replacement, then the replacement is accepted with probability $(1 - s_i)$. Otherwise, no replacement occurs and the original individual does not die.

1.2 Multidimensional diffusion limit

Assume that the relative fitnesses and mutation rates scale as

$$\frac{\phi_{ij}}{N} = \beta_{ij}, \qquad \frac{s_i}{N} = \gamma_i.$$

The infinitesimal drift and variance are now a vector and matrix respectively. For instance the expected displacement of the frequency of i-types started at

$$x = (x_1, \ldots, x_K) = \left(\frac{n_1}{2N}, \ldots, \frac{n_K}{2N} \right)$$

is

$$\begin{aligned} &\mathbb{E}_x[\Delta_h X_i(0)] \\ &\quad = \frac{h}{2N} \left\{ \sum_{j \neq i} n_j \left(\frac{n_i}{2N}(1 - s_i) + \phi_{ji} \right) - \sum_{j \neq i} n_i \left(\frac{n_j}{2N}(1 - s_j) + \phi_{ij} \right) \right\} + o(h). \end{aligned}$$

Running time at rate N and taking a limit $N \to \infty$, we get the infinitesimal drift

$$\mu_i(x) = -x_i \sum_j \beta_{ij} + \sum_j x_j \beta_{ji} + x_i \sum_j x_j(\gamma_j - \gamma_i).$$

(We set $\phi_{ii} = 0$.) Similarly, for the covariance,

$$\begin{aligned} &\mathbb{E}_x[\Delta_h X_i(0) \Delta_h X_j(0)] \\ &\quad = \frac{h}{(2N)^2} \left\{ n_j \left(\frac{n_i}{2N}(1 - s_i) + \phi_{ji} \right) + n_i \left(\frac{n_j}{2N}(1 - s_j) + \phi_{ij} \right) \right\} + o(h). \end{aligned}$$

Running time at rate N and taking a limit $N \to \infty$, we get the infinitesimal covariance,

$$\sigma_{ij}^2(x) = -x_i x_j.$$

Similarly,

$$\sigma_{ii}^2(x) = x_i(1 - x_i).$$

2 Hill-Robertson interference

As an application of the previous section, we consider the interference between two advantageous alleles. Consider two loci with alleles A/a and B/b. Suppose that the population is originally made of ab, and that the advantageous alleles A and B arise. If A arises first its genetic background is Ab. If, further, B arises when A is still at a low frequency it is likely that the genetic background of the B mutation is aB. The *Hill-Roberston interference* is the observation that, in the absence of recombination between the two loci, there is a competition between the fixation of A and B. The overall fixation probability of either allele is then reduced. This is an argument for the evolution of recombination.

Formalization. Suppose we have three alleles aB, Ab, and ab, which we call 1, 2, and 3 respectively, with relative fitnesses $1 - s$, $1 - s$ and $1 - 2s$. To compute the probability that 1 fixates in the absence of mutation, $u(x)$, we need to solve (as we did in the one-dimensional case)

$$Lu = 0,$$

with appropriate boundary conditions, where the infinitesimal generator is

$$\begin{aligned} Lf = \frac{1}{2}x_1(1 - x_1)D_{11}f - x_1x_2D_{12}f + \frac{1}{2}x_2(1 - x_2)D_{22}f \\ + x_1 \sum_j x_j(\gamma_j - \gamma_1)D_1 f + x_2 \sum_j x_j(\gamma_j - \gamma_2)D_2 f, \end{aligned}$$

where we only keep track of $x_1 \geq 0$ and $x_2 \geq 0$ with domain $0 \leq x_1 + x_2 \leq 1$. (For a formal justification, see [Dur08].) The boundary conditions can easily be derived from the one-dimensional case. When $x_1 = 0$, $u(x) = 0$. When $x_1 + x_2 = 1$, $u(x) = x_1$ (since this is effectively the same as a neutral case). When $x_2 = 0$,

$$u(x) = \frac{1 - e^{-2\sigma x_1}}{1 - e^{-2\sigma}},$$

where $\sigma = \gamma_3 - \gamma_1 = 2Ns$.

It can be checked [Dur08] that the solution is

$$u(x) = \frac{1 - e^{-2\sigma(x_1+x_2)}}{1 - e^{-2\sigma}} \frac{x_1}{x_1 + x_2}.$$

The answer is somewhat intuitive as it corresponds to 3 losing against the combined $1/2$ followed by a fair game between 1 and 2. For x_1, x_2 small, we have the expansion,

$$u(x) \approx \frac{2\sigma x_1 - 2\sigma^2 x_1(x_1 + x_2)}{1 - e^{-2\sigma}}.$$

In comparison, in the absence of allele 2,

$$u(x) \approx \frac{2\sigma x_1 - 2\sigma^2 x_1^2}{1 - e^{-2\sigma}}.$$

If, instead, there was recombination between the two loci, the allele AB could be formed and both alleles could fixate.

Further reading

The material in this section was taken from Chapter 8 of the excellent monograph [Dur08].

References

[Dur08] Richard Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer, New York, second edition, 2008.